# Towards integrated models of regulatory networks: the MetaGenoReg project

Hidde de Jong and Daniel Kahn

INRIA Grenoble - Rhône-Alpes

Hidde.de-Jong@inria.fr

http://ibis.inrialpes.fr

LBBE, Université de Lyon Daniel.Kahn@univ-lyon1.fr http://lbbe.univ-lyon1.fr



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



centre de recherche GRENOBLE - RHÔNE-ALPES





### Overview

- General question of biological regulation: the MetaGenoReg project
- 2. Biological system: the glucose-acetate diauxie in *E. coli*
- 3. Methodological approach
- 4. Metabolic model
- 5. Integrating gene and metabolic models







### 1. General question of biological regulation

#### Cellular regulation involves several levels, including:

- Gene regulatory networks
- Metabolic regulation
- These levels interact:
  - Gene expression impacts metabolism through changes in enzyme concentrations
  - Conversely metabolism influences gene expression
- What is the rationale articulating both types of regulation?
  - Are they interchangeable ?
  - How much are they constrained?
  - What is the relative importance of gene and metabolic regulation?







### 'Hierarchical' analysis



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



centre de recherche GRENOBLE - RHÔNE-ALPES





4

### MetaGenoReg project outline

Modelling combined metabolic and gene regulation

- Reduce and simplify in order to understand the system's behaviour
- Develop a method for joint modelling combining different approximations suited to both types of regulation
- Measure their respective contribution
- Analyse the model's strengths and weaknesses from a systemic point of view
- Understand the biological rationale underlying the distribution of regulation between metabolism and gene expression









### Glucose-acetate diauxie

- Well-characterised transition in *E. coli*
- Involves major changes
  - at the metabolic level: gluconeogenesis vs. glycolysis
  - at the gene expression level
- Strong interaction between metabolic and gene expression levels



Oh et al. (2002), J Biol Chem. 277(15):13175-83.







### Carbon assimilation in *E. coli*

Interactions between metabolism and gene expression involve complex regulatory networks

Genetic and metabolic control of glycolysis and gluconeogenesis



### 3. Methodological approach

- Development of integrated model of upper-part of carbon assimilation in *E. coli*
- Kinetic model consisting of 41 variables and more than 100 parameters
- Main problems with model: lack of parameter values, lack of data to estimate parameter values
- Basic assumption for model reduction: on the time-scale of gene expression, metabolism is a fast process

In bacteria, time constants for gene expression are typically of the order of (tens of) minutes, whereas time constants in metabolism are typically of the order of seconds











- How is time-hierarchy exploited to formally reduce kinetic model of integrated genetic and metabolic network?
- **\*** Basic form of kinetic model  $\dot{x} = N v(x)$ 
  - Concentration variables  $x \in \mathbb{R}^n_+$
  - Reaction rates  $v\,:\,\mathbb{R}^n_+ o\mathbb{R}^q$
  - Stoichiometric matrix  $N \in \mathbb{Z}^{n imes q}$
- ★ Time-scale hierarchy motivates distinction between fast reaction rates  $v^f \in \mathbb{R}^{q-p}$  and slow reaction rates  $v^s \in \mathbb{R}^p$  such that  $v = [v^s \ v^f]'$

Typically, enzymatic and complex formation reactions are fast, protein synthesis and degradation are slow







✤ Separation of fast and slow reactions motivates a linear transformation  $T \in \mathbb{Z}^n \times \mathbb{Z}^n$  of the variables

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = T x \quad \text{such that} \quad \begin{bmatrix} N^s & 0 \\ N^{s'} & N^f \end{bmatrix} = T N$$

♦ We call  $x^s \in \mathbb{R}^m_+$  slow variables and  $x^f \in \mathbb{R}^{n-m}_+$  fast variables, while  $N^s \in \mathbb{Z}^m \times \mathbb{Z}^p$  and  $N^{s'} \in \mathbb{Z}^{n-m} \times \mathbb{Z}^p$  are stoichiometric matrices for slow reactions and  $N^f \in \mathbb{Z}^{n-m} \times \mathbb{Z}^{q-p}$ is stoichiometric matrix for fast reactions

Slow variables are typically **total protein concentrations**, fast variables **metabolites and biochemical complexes** 









Separation of fast and slow variables allows original model to be rewritten as coupled slow and fast subsystems

$$\dot{x}^{s} = N^{s} v^{s}(x^{s}, x^{f})$$
$$\dot{x}^{f} = N^{s'} v^{s}(x^{s}, x^{f}) + N^{f} v^{f}(x^{s}, x^{f}) \approx N^{f} v^{f}(x^{s}, x^{f})$$

Under quasi-steady-state approximation (QSSA), fast variables are assumed to instantly adapt to slow dynamics

$$\dot{x}^f = 0 \implies N^f v^f(x^s, x^f) = 0$$

Mathematical basis for QSSA is given by Tikhonov's theorem

Heinrich and Schuster (1996), *The Regulation of Cellular Systems*, Chapman & Hall Khalil (2001), *Nonlinear Systems*, Prentice Hall, 3rd ed.







QSSA implicitly relates steady-state value of fast variables to slow variables

 $x^f = g(x^s), g : \mathbb{R}^m_+ \to \mathbb{R}^{n-m}_+$ 

#### This gives reduced model on the slow time-scale

 $\dot{x}^s = N^s v^s(x^s, q(x^s))$ 

Reduced model describes direct and indirect dependencies between slow variables (total protein concentrations)

Mathematical representation of effective gene regulatory network

- Notice
  - Generally function  $\underline{q}$  is not easy to obtain due to nonlinearities
  - Function q depends on unknown parameter values









### Jacobian matrix and regulatory structure

Derivation of interaction structure between slow variables by computation of **Jacobian matrix** 

$$\mathcal{J} = \frac{\partial \dot{x}^s}{\partial x^s} = N^s \frac{\partial v^s(x^s, x^f)}{\partial x^s} + N^s \frac{\partial v^s(x^s, x^f)}{\partial x^f} \frac{\partial g(x^s)}{\partial x^s}$$
Direct regulation by
transcription factors
Indirect regulation
through metabolism

• Implicit differentiation of  $N^f v^f(x^s, x^f) = 0$  yields

$$\frac{\partial g(x^s)}{\partial x^s} = \underbrace{-M^{-1} N^f}_{\gamma} \frac{\partial v^f(x^s, x^f)}{\partial x^s}$$

**Concentration control coefficients** 

where  $M = N^f \partial v^f (x^f, x^s) / \partial x^f$  is Jacobian matrix of fast system









### **Determination of interaction signs**

- Can we derive signs for regulatory interactions (elements of Jacobian matrix), without knowledge on rate laws and parameter values?
- Idea: exploit fact that signs of elasticities are known

Rate laws are generally monotone functions in variables









### Determination of interaction signs

- Can we derive signs for regulatory interactions (elements of Jacobian matrix), without knowledge on rate laws and parameter values?
- Idea: exploit fact that signs of elasticities are known

Rate laws are generally monotone functions in variables

- Notice
  - Reversible reactions: signs of  $\partial v^f(x^s,x^f)/\partial x^s$  change with flux direction

$$\begin{array}{ccc} (x^s) & \mathrm{E} & & \\ & & \mathbf{m}_1 & \stackrel{\nabla}{\xrightarrow{}} & \mathbf{m}_2 & & & \frac{\partial v^f}{\partial x^s} > 0 \end{array}$$







### Determination of interaction signs

Resolution of signs of (large) algebraic expressions defining interaction signs by means of computer algebra tools

$$\mathcal{J} = \frac{\partial \dot{x}^s}{\partial x^s} = N^s \, \frac{\partial v^s(x^s, x^f)}{\partial x^s} + N^s \, \frac{\partial v^s(x^s, x^f)}{\partial x^f} \, \frac{\partial g(x^s)}{\partial x^s}$$

Symbolic Math Toolbox in Matlab

Use of additional constraints in sign resolution

- Stability assumption for fast system: necessary condition for stability is that coefficients of characteristic polynomial  $det(M \lambda I) = 0$  have same sign
- Experimental determination of some of the signs of concentration control coefficients in  $\frac{\partial g(x^s)}{\partial x^s}$  (if available)







### Application to E. coli carbon assimilation

Development of model of carbon assimilation network, analysis under following conditions:

Glycolysis/gluconeogenesis (growth on glucose/pyruvate)



Baldazzi et al. (2010), PLoS Comput. Biol., 6(6):e1000812









### Application to E. coli carbon assimilation

Development of model of carbon assimilation network, analysis under following conditions:

Glycolysis/gluconeogenesis (growth on glucose/pyruvate)



0	1+	5	1.00	-					Gyind	Gyn	торя	Rpos	RSSB	SLADIE RIVAS	1101
0	-1+				-	0	0	0	0	0	0	0	0	0	-
	-/ •	-/+	-/+	-/+		+	+	0	0	0	0	0	0	0	
0	-/+	-/+	-/+	-/+	100	+	+	0	0	0	0	0	0	0	
0	-/+	-/+	-/+	-/+	-	+	+	0	0	0	0	0	0	0	-
0	-	-	-	-	-	0	0	0	0	0	0	0	0	0	-
0	-	-	-	-	-	0	0	0	0	0	0	0	0	0	-
0	-	-	-	-	+	-	-	0	0	0	0	0	0	0	0
0	+	+	+	+	-	+	+	-	0	0	0	0	0	0	0
0	0	0	0	0	0	-	-	-	+	-	-	0	0	0	0
0	0	0	0	0	0	0	0	-	-	+	+	0	0	0	0
0	0	0	0	0	0	+	+	0	0	0	0	+	0	0	0
0	0	0	0	0	0	0	0	+	+	-	-	+	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0
0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0
0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0
0	-	-	-	-	-	0	0	0	0	0	0	0	0	0	-

Regulators

 $\mathcal{J} = \frac{\partial \dot{x}^s}{\partial x^s}$ 

**Glycolysis with allosteric effects** 

Few fast variables couple metabolism to gene expression







### Network is densely connected

- Contrary to what is often maintained, gene regulatory network is found to be densely connected
- Strong connectivity arises from indirect interactions mediated by metabolism
  - $\mathcal{M}^0$ : transcriptional network consisting of direct interactions only
  - $\mathcal{M}^2_{glyco}$  : gene regulatory network in glycolytic growth conditions including direct and indirect interactions

	$\mathcal{M}^{0}$	$\mathcal{M}^1_{glyco}$	${\cal M}^2_{glyco}$	$\mathcal{M}^1_{neo}$	$\mathcal{M}^2_{neo}$
Number of feedback loops	4	2388	9246	24	2257
Maximal loop length	2	12	12	6	12
Average connectivity	1.4	4.7	5.2	2.8	4.4

#### Experimental evidence for indirect interactions

Siddiquee et al. (2004), FEMS Microbiol. Lett., 235:25-33











### Network is largely sign-determined

## Derived gene regulatory network for carbon assimilation in *E. coli* is largely sign-determined

Signs of interactions do not depend on explicit specification of kinetic rate laws or parameter values, but are structural property of system

	regulators																
		PfkA	FbaA	GapA	Pgk	Eno	PykF	Cya	Crp	Fis	GyrAB	Gyrl	TopA	RpoS	RssB	stable RNAs	FruR
	pfkA	0		-	-	-	182	0	0	0	0	0	0	0	0	0	-
	fbaA	0	-/+	-/+	-/+	-/+		+	+	0	0	0	0	0	0	0	
	gapA	0	-/+	-/+	-/+	-/+		+	+	0	0	0	0	0	0	0	
	pgk	0	-/+	-/+	-/+	-/+	-	+	+	0	0	0	0	0	0	0	-
	eno	0	-6	(e)	-	- 60	-	0	0	0	0	0	0	0	0	0	140
	pykF	0	-2	-	-	20		0	0	0	0	0	0	0	0	0	1.00
	cya	0	-	-	-	-	+	2	-	0	0	0	0	0	0	0	0
Genes	crp	0	+	+	+	+		+	+		0	0	0	0	0	0	0
	fis	0	0	0	0	0	0	1	12	-	+		2	0	0	0	0
	gyrAB	0	0	0	0	0	0	0	0	127		+	+	0	0	0	0
	gyrl	0	0	0	0	0	0	+	+	0	0	0	0	+	0	0	0
	topA	0	0	0	0	0	0	0	0	+	+	120		+	0	0	0
Chucohusis with alloctaria offects	rpoS	0	0	0	0	0	0	0	0	0	0	0	0	0	( <b>5</b>	0	0
Giveniysis with anosteric effects	rssB	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0
	rrn	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0
	fruR	0	-0	-	-			0	0	0	0	0	0	0	0	0	

Sign-determinedness not expected on basis of work in ecology

Sufficient conditions for sign-determinedness can be formulated using

Baldazzi et al. (2010), PLoS Comput. Biol., 6(6):e1000812

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



expression for  $\mathcal{J}$ 



Demulator





### Interaction signs change with fluxes

Radical changes in environment may invert signs of indirect interactions, because they change direction of metabolic fluxes and thus signs of elasticities





Network under glycolytic conditions

Network under gluconeogenic conditions

Dynamic modification of feedback structure in response to environmental perturbations







### Key findings and new questions

Systematic derivation of effective structure of gene regulatory network on time-scale of gene expression

Weak assumptions on time-scale hierarchies and stability

Obtained network is at the same time robust and flexible

- Robust to changes of kinetic properties (results not dependent on parameter values and rate laws)
- Flexible rewiring of network structure following radical changes in environment (changes in flux directions)
- Results on *E. coli* network raise several issues:
  - To which extent do observations carry over to other regulatory systems in bacteria and higher organisms?
  - How do indirect interactions affect dynamics of networks?







### Dynamical analysis of networks

- Reduced networks describe direct and indirect regulatory interactions between genes
- Qualitative dynamics of gene regulatory interactions can be described by PL models
   Glass and Kauffman (1973), J. Theor. Biol., 39(1):103-29
- Translation of network diagrams into PL models
  - Straightforward for direct interactions...
  - ... but also possible for indirect interactions





INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE







24

### Qualitative analysis of PL models



#### PL models using step functions



#### Model-checking for verification of system properties

de Jong et al. (2004), Bull. Math. Biol., 66(2):301-40 Batt et al. (2007), Automatica, 44(4):982-89

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE AUTOMATIQUE





 $max_h$ 

 $K_b/\gamma_b$ 

 $\theta_{\rm h}$ 

0





25



Predictions of qualitative dynamics, robust for large variations in parameter values

Models easy to analyze, using inequalities  $D^{20}$  $D_{\bullet}^{25}$ 

 $\theta_{a2}$ 

max<sub>a</sub>

 $\theta_{a1}$ 

### Development of PL model

PL models for growth on glucose and growth on acetate

- Models have13 equations and require specification of few dozen parameter inequalities
- Calibration of PL models by means of qualitative data
  - Literature data
  - Extension of previous model of network of global regulators

Ropers et al. (2006), Biosystems, 84(2):124-152; Ropers et al. (2009), in press

• Hypotheses (educated guesses)







### Aim of analysis of PL model

#### Aims of analysis of network dynamics:

- Predict evolution of gene expression levels after diauxic shift
- Study role of indirect interactions mediated by metabolism

#### Modeling of batch experiments

- Grow bacteria in M9 with glucose to steady state (glycolysis)
- Continue growth on (excreted) acetate after glucose exhaustion (gluconeogenesis)

Brice Enjalbert, personal communication







### Some preliminary results

- Cross-inhibition between Fis and Crp predicted to play central role in adaptation of gene expression upon glucose depletion
- Predicted gene expression profiles verified by means of reporter gene measurements









### Towards quantitative models?

- Above approach leads to models that view metabolism as intermediary between gene regulatory interactions
- However, metabolism is not explicitly modeled

PL models aggregate and approximate complex rate functions in reduced model

$$\dot{x}^s = N^s v^s(x^s, g(x^s)) \Rightarrow \dot{x}^s = f(x^s)$$

Moreover, models provide qualitative instead of quantitative picture of dynamics

Qualitative models help provide intuitive idea of global system dynamics, but for some questions quantitative precision is required









### Towards quantitative models?

Another approach explicitly models metabolism and gene expression, followed by integration of two parts

$$N^{f} v^{f}(x^{s}, x^{f}) = 0 \implies x^{f} = g(x^{s})$$
$$\dot{x}^{s} = N^{s} v^{s}(x^{s}, x^{f})$$

- $\clubsuit$  Approach based on suitable approximations of g
  - Approximations should provide good phenomenological description of metabolic rate laws
  - Minimal number of parameters to facilitate identification of parameter values from experimental data







### 4. Metabolic model

- Toy model entirely specified with ODEs
- 'Experimental' object used to test the quality of various reductions and approximations by comparison of simplified models with complete ODE model
- A suitable approximation would ideally allow us to calculate  $g(x^s)$  analytically







### Which approximations?

- Various types of linearisation of metabolic effects
- Compare reduced / approximated models with complete ODE-specified model







### Assessing approximations of metabolism

- Randomly change enzyme concentrations in a 25-fold range on benchmark model (Matteo Brilli)
- Test steady-state obtention



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



GRENOBLE - RHÔNE-ALPES





### Approximation 1, from MCT

$$\mathbf{v} = diag \mathbf{f}(\mathbf{x}) \cdot \mathbf{e}$$

#### $d\mathbf{X}/d\mathbf{e} = \Gamma \cdot d\mathbf{v}/d\mathbf{e}$

$$\Delta \ln \mathbf{X} \sim (diag \ \mathbf{X}_0)^{-1} \cdot \Gamma \cdot diag \ \mathbf{f}(\mathbf{X}_0) \cdot \Delta \mathbf{e}$$

#### Linearization around steady-state using control coefficients









### Approximation 2, linlog

Linearization of kinetic laws:

$$\mathbf{v}(\mathbf{x}) \sim diag \mathbf{e} \cdot (\mathbf{A} + \mathbf{B} \cdot \ln \mathbf{x})$$

Steady-state implies:

 $\mathbf{N} \cdot \mathbf{v}(\mathbf{X}) = \mathbf{0}$ 

$$\ln \mathbf{X} \sim -(\mathbf{N} \cdot diag \mathbf{e} \cdot \mathbf{B})^{-1} \cdot \mathbf{N} \cdot diag \mathbf{e} \cdot \mathbf{A}$$







### Approximation 3, hyperbolic

Suggested from earlier work by Kacser:

$$\Delta(1/\mathbf{X}) \sim (diag \mathbf{X}_0)^{-1} \cdot \mathbf{C}^X \cdot diag \mathbf{e}_0 \cdot \Delta(1/\mathbf{e})$$

$$\Delta(1/\mathbf{J}) \sim (diag \mathbf{J}_0)^{-1} \cdot \mathbf{C}^J \cdot diag \mathbf{e}_0 \cdot \Delta(1/\mathbf{e})$$

Linearization around steady-state using control coefficients







### Metabolite estimates

Root mean square Log deviation



Approximation MCA

Approximation linlog

Approximation hyperbolic

### Flux estimates

#### Median flux absolute Log deviation



Approximation MCA

Approximation linlog

Approximation hyperbolic

### Model of *E. coli* carbon metabolism

- Simplified model
  - 32 reactions
  - 17 metabolites
- Linlog approximation
  J ~ diag e · (A + B · lnX)
- → Independent linear regression possible for each reaction if sufficient data available :
  - Fluxes
  - Enzyme expression
  - Metabolite concentrations









### Issues with metabolic model identification

- Difficulties to obtain high quality complete datasets (fluxes, metabolite <u>and</u> enzyme concentrations) with sufficient numbers of distinct observations
- Missing data can be handled efficiently by EM or maximum likelihood methods (Berthoumieux *et al.*, submitted)
- Identifiability issues arise when there is insufficient variability or dependencies between metabolite concentrations because of
  - Reactions close to equilibrium
  - Steady-state constraints
  - Homeostasis
  - → Usefulness of dynamic non steady-state measurements (difficult to obtain)







### Working around identifiability issues

Use Principal Component Analysis on InX

• Reduce metabolite data by Singular Value Decomposition

 $\ln \mathbf{X} - \overline{\ln \mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{V}^{T}$ 

- Determine effective dimension of  $\ln X$  from singular values  $\sigma_i$ , neglecting  $\sigma_i^2$  smaller than experimental variance
- Reduce metabolite data and reformulate identification accordingly

### $\mathbf{Y} = \mathbf{U}_r^T \ln \mathbf{X}$

- Estimate parameters  $\mathbf{B}_r$  for the reduced model such that

$$\mathbf{J} / \mathbf{e} - \overline{\mathbf{J} / \mathbf{e}} = \mathbf{B} \ln \mathbf{X} = \mathbf{B}_r \mathbf{Y}$$

• One among an infinite number of choices for full parameters is  $\mathbf{B} = \mathbf{B}_r \mathbf{U}_r^T$ 







### 5. Integrating gene and metabolic models

- Identify separately the fast component (metabolic) and the slow component (gene expression)
- Use the resulting analytical model of metabolic steady-states as a 'plugin' function in the gene network model

$$N^{f} v^{f}(x^{s}, x^{f}) = 0 \implies x^{f} = g(x^{s})$$
$$\dot{x}^{s} = N^{s} v^{s}(x^{s}, x^{f})$$

Critical issue: identification methods and, especially, quality and quantity of experimental data







### Experimental data on metabolism

Quantification of extra-cellular metabolites by means of nuclear magnetic resonance (NMR) spectroscopy

Brice Enjalbert, personal communication







### Experimental data on metabolism

Quantification of intra-cellular metabolites by means of mass spectrometry

Brice Enjalbert, personal communication







### **Isotope Dilution Mass Spectrometry**

### Experimental data on metabolism

Quantification of intra-cellular metabolites by means of mass spectrometry

Brice Enjalbert, personal communication







### Experimental data on gene expression

- Quantification of gene expression by means of fluorescent and luminescent reporter genes
  - Expression of reporter gene is proportional to expression of target gene











### Experimental data on gene expression



Hans Geiselmann, personal communication

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE



GRENOBLE - RHÔNE-ALPES





### **Prospect:**

### Roles of metabolic and gene regulation

- Identify parameters of the reduced system from data
- Study the metabolic response in the model when gene regulation is abolished
- Evaluate (quantify) the contribution of gene regulation to the metabolic response
- Conversely calculate the contribution of metabolic effects to gene regulation
- Understand the biological rationale underlying the distribution of regulation between metabolism and gene expression







49